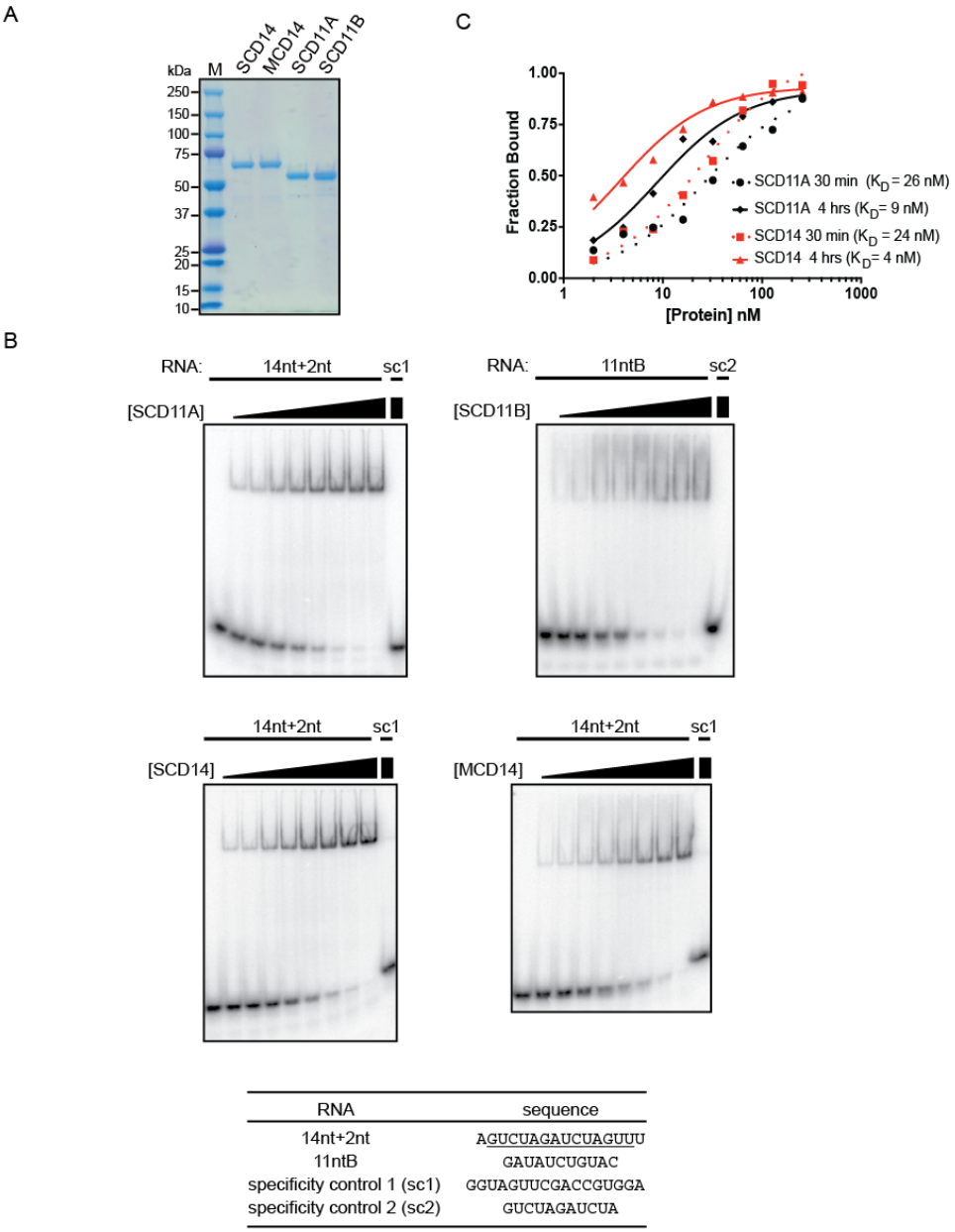
Supplemental Information, Miranda et al.

Table S1. Effects of single nucleotide substitutions on the predicted stability of RNA structure in the RNA ligand of SCD14 and MCD14. The binding site for which these proteins were designed is shown at the top. Single nucleotide changes at each position are highlighted in red. The predicted ΔG of RNA structure as predicted by Mfold (1) is shown to the right.

Sequence	Position of Substitution	RNA 2° Structure ΔG (kcal/mol)
GUCUAGAUCUAGUU	-	0.3
AUCUAGAUCUAGUU	1	0.3
CUCUAGAUCUAGUU	1	0.3
UUCUAGAUCUAGUU	1	0.3
GACUAGAUCUAGUU	2	**-2**
GGCUAGAUCUAGUU	2	**-1.5**
GCCUAGAUCUAGUU	2	0.1
GUAUAGAUCUAGUU	3	2.5
GUGUAGAUCUAGUU	3	2.6
GUUUAGAUCUAGUU	3	2.2
GUCAAGAUCUAGUU	4	3.4
GUCGAGAUCUAGUU	4	3.4
GUCCAGAUCUAGUU	4	3.4
GUCUGGAUCUAGUU	5	0.7
GUCUCGAUCUAGUU	5	2.6
GUCUUGAUCUAGUU	5	1.8
GUCUAAAUCUAGUU	6	1
GUCUACAUCUAGUU	6	1.3
GUCUAUAUCUAGUU	6	1.2
GUCUAGGUCUAGUU	7	0.3
GUCUAGCUCUAGUU	7	0.3
GUCUAGUUCUAGUU	7	0.3
GUCUAGAACUAGUU	8	0.3
GUCUAGAGCUAGUU	8	0.3
GUCUAGACCUAGUU	8	0.3
GUCUAGAUAUAGUU	9	0.4
GUCUAGAUGUAGUU	9	1.3
GUCUAGAUUUAGUU	9	1.7
GUCUAGAUCAAGUU	10	2.1
GUCUAGAUCGAGUU	10	1.1
GUCUAGAUCCAGUU	10	2.3
GUCUAGAUCUGGUU	11	0.6
GUCUAGAUCUCGUU	11	3.4
GUCUAGAUCUUGUU	11	3.4
GUCUAGAUCUAAUU	12	2.7
GUCUAGAUCUACUU	12	2.2
GUCUAGAUCUAUUU	12	2.9
GUCUAGAUCUAGAU	13	**-2.3**
GUCUAGAUCUAGGU	13	**-0.8**
GUCUAGAUCUAGCU	13	0.2
GUCUAGAUCUAGUA	14	0.3
GUCUAGAUCUAGUG	14	0.3
GUCUAGAUCUAGUC	14	0.3

Supplementary Table S1

Supplementary Figure S1. Purity and RNA binding activity of designer PPR proteins used in this study. **(A)** SDS-PAGE of purified proteins stained with Coomassie Blue. **(B)** Gel mobility shift assays demonstrating affinity of each protein for specific and non-specific RNAs. The sequences of the specific RNAs and of specificity controls (sc) are shown below. Proteins were assayed at 0, 2, 4, 8, 16, 32, 64, 128, and 256 nM. **(C)** Comparison of apparent affinities of SCD14 and SCD11A for their specific RNA ligand (14nt+2nt RNA, see **B**) based on binding reactions incubated for 30 min or 4 hr. Binding was monitored with gel mobility shift assays. These assays used fresh protein preparations, whereas all other experiments used preparations that had been stored at -20˚C for several months.

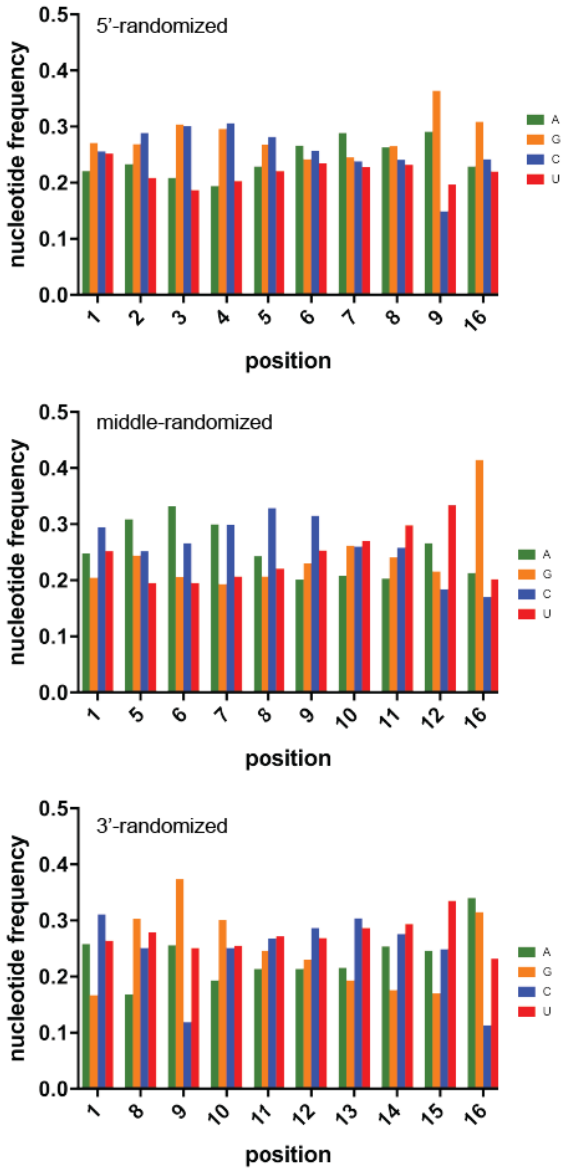


RNA	sequence
14nt+2nt	AGUCUAGAUCUAGUUU
11ntB	GAUAUCUGUAC
specificity control 1 (sc1)	GGUAGUUCGACCGUGGA
specificity control 2 (sc2)	GUCUAGAUCUA

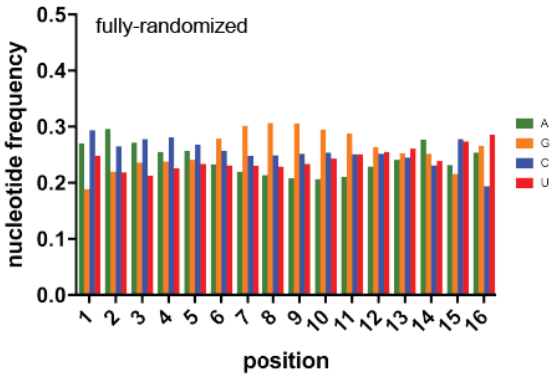
Figure S1

Supplementary Figure S2. Characteristics of the randomized RNA pools used for bind-n-seq assays. **(A)** Nucleotide frequencies at each position in each partially randomized RNA pool. **(B)** Nucleotide frequencies at each position in the fully randomized RNA pool. **(C)** Number of sequence reads obtained for input fractions, and coverage of anticipated sequence space in the sequenced input aliquots. Each partially randomized RNA is expected to include 4^{10} (1,048,576) different sequences. Slightly less than this number was detected in the aliquots that were sequenced. The fully randomized 16-mer is expected to include 4^{16} different sequences, a number that is impractical to thoroughly sample by sequencing. The informatics approaches used for experiments involving the fully-randomized RNA did not rely on a comprehensive analysis of the input sequence population. Roughly 500,000 reads were obtained for each bound fraction.

A



B

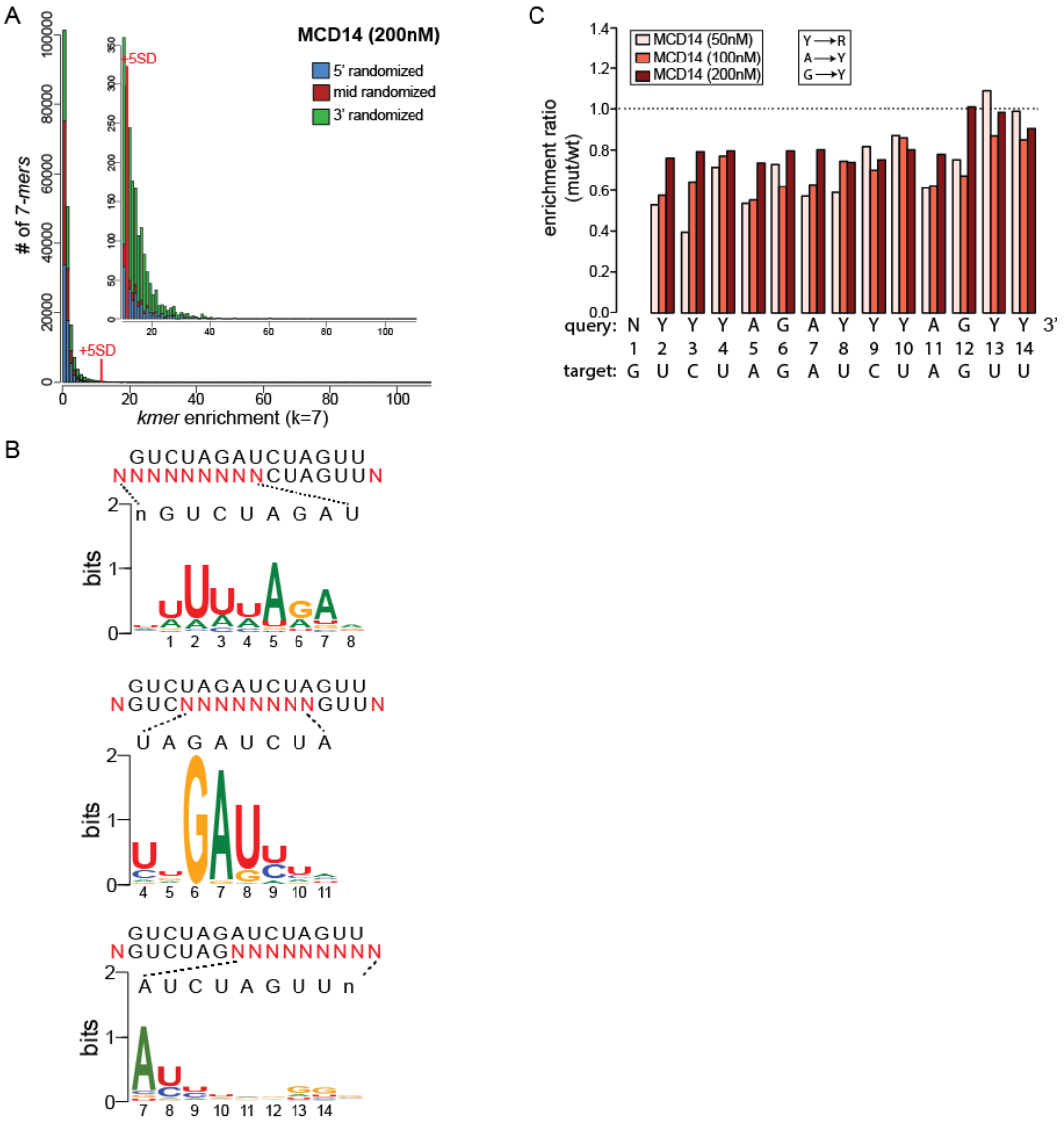


C

oligonucleotide	reads	distinct sequences
5' randomized	18,526,067	998,243
mid randomized	31,347,239	1,003,052
3' randomized	39,165,618	991,102
fully randomized	26,868,915	23,828,931

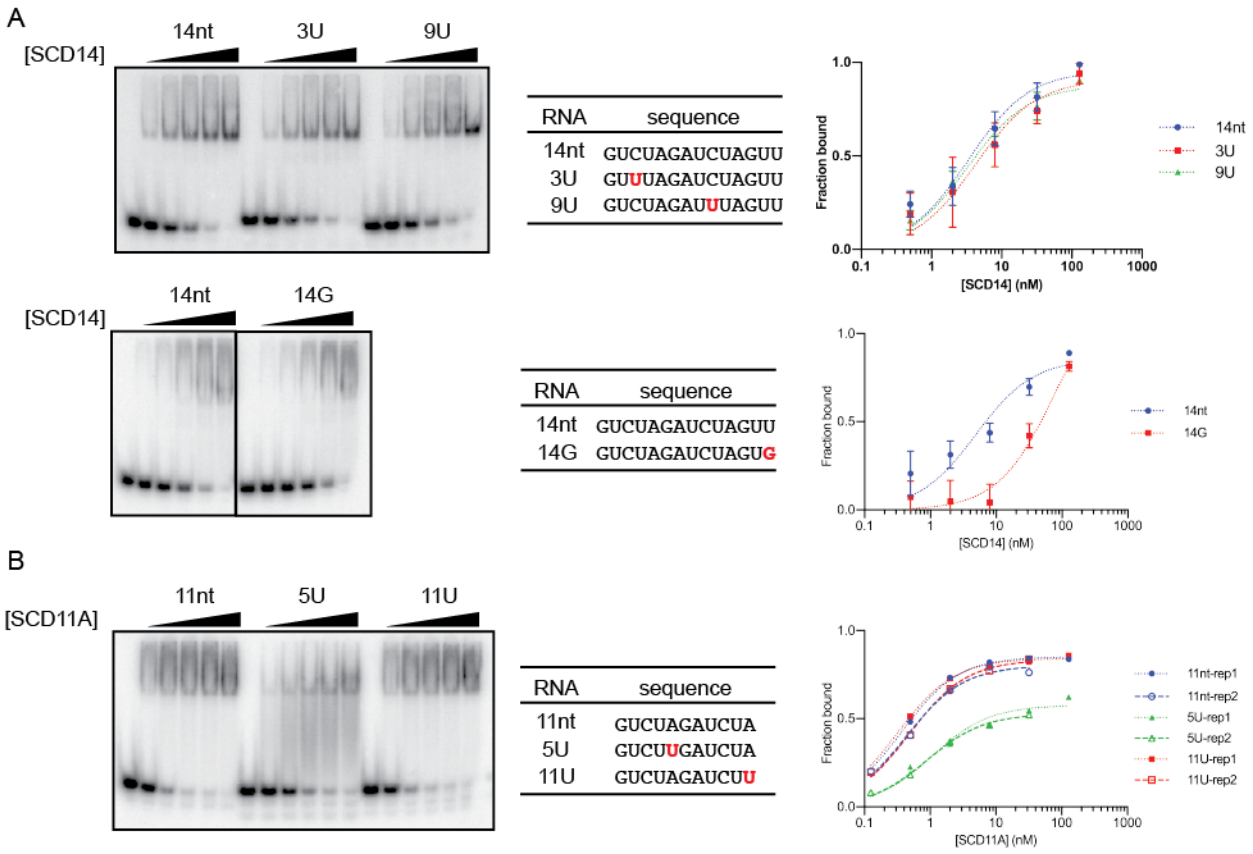
Supplementary Figure S2

Supplementary Figure S3. Analyses of Bind-n-Seq data for MCD14. **(A)** Frequency distribution of enrichment values of all 7-mers for the bind-n-seq assay using MCD14 (200 nM) and equimolar amounts of the partially randomized RNA pools. The graph shows the number of different 7-mers (Y-axis) at each enrichment value (X-axis). The inset shows an expansion of the tail of the distribution. 7-mers that were enriched more than 5 standard deviations above the mean were defined as the highly enriched fraction for subsequent analyses. **(B)** Sequence logos derived from sequences harboring highly enriched 7-mers for the bind-n-seq assay using MCD14 (200 nM) and equimolar amounts of the three partially randomized RNA pools. The randomized nucleotides (N) are displayed beneath the intended binding site. Sequences harboring highly enriched 7-mers (>5 SD above the mean) were used to generate each logo, after weighting the input sequences according to their enrichment values. The logos from the 5'-randomized, middle-randomized, and 3'-randomized RNAs were calculated from 3743 sequences harboring 231 7-mers, 1224 sequences containing 59 7-mers, and 13545 sequences harboring 1019 7-mers, respectively. **(C)** Tolerance of MCD14 to transversion mismatches at each position. The analysis is the same as that described in Figure 4A for SCD14.



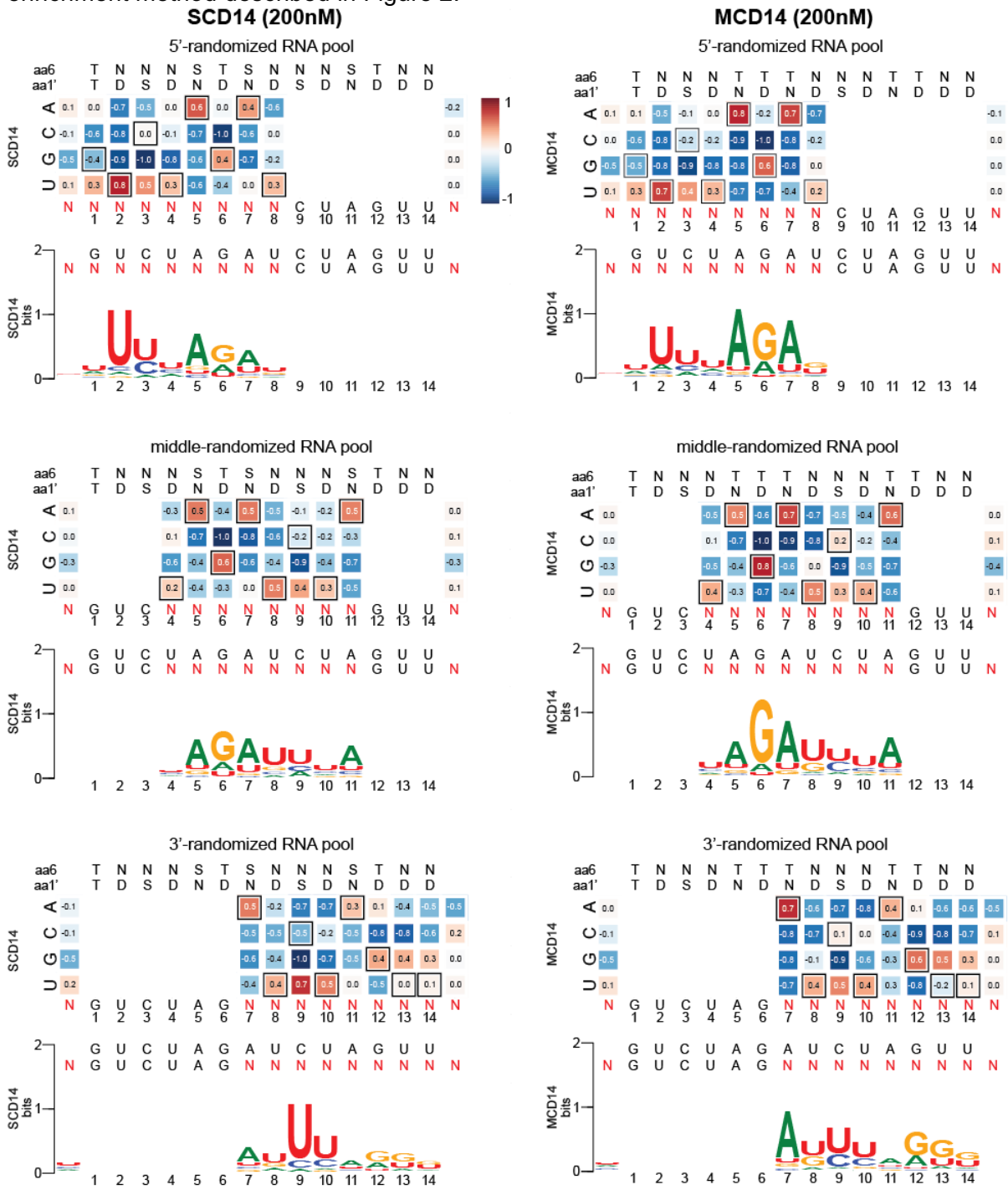
Supplementary Figure S3

Supplementary Figure S4. Gel mobility shift assays testing inferences from the bind-n-seq data. Proteins were assayed at 0, 0.5, 2, 8, 32, and 128 nM. **(A)** Assays of substitutions in the SCD14 binding site. Each experiment was repeated three times. Raw data for representative assays are shown. Median values and standard deviations are shown in the plots. The two images separated by a line in the 14G experiment came from non-contiguous lanes on the same gel. **(B)** Assays of substitutions in the SCD11A binding site. The results confirm that an A-to-U transversion mismatch at the center of the binding site causes a much stronger loss of binding affinity than does an A-to-U transversion at the 3' end. Each experiment was repeated twice and data from both replicates are shown in the graph.



Supplementary Figure S4

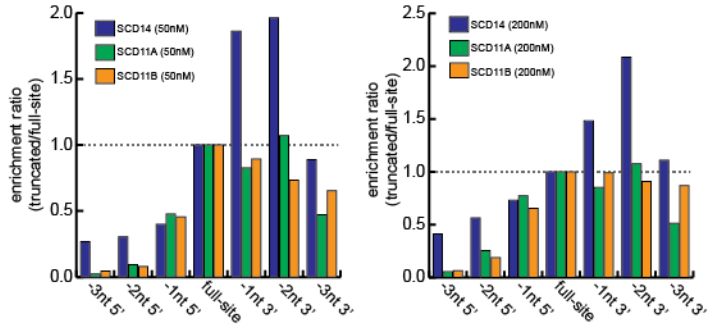
Supplementary Figure S5. Nucleotide selectivity of each PPR motif in SCD14 and MCD14. Specificity scores for each nucleotide position were calculated from highly enriched sequences (> 5 SD above the mean) in bind-n-seq assays with each protein (200 nM) and either the 5', middle, or 3' randomized RNA pool. Scores were calculated as in (2). Positive and negative values indicate selection for or against the indicated nucleotide, respectively. The nucleotide predicted by the PPR code at each position is marked with a black box. The logos below were generated using the *7*-mer enrichment method described in Figure 2.



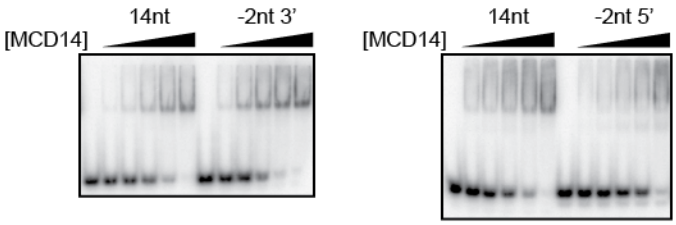
Supplementary Figure S5

Supplementary Figure S6. Effects of binding site truncation on sequence enrichment.
(A) The analysis is the same as that in Figure 5A except that the data came from experiments with different protein concentrations. **(B)** Gel mobility shift analysis confirming the effects of binding site truncations on affinity for MCD14 *in vitro.* The analysis is the same as that in Figure 5B except that MCD14 was used rather than SCD14.

A



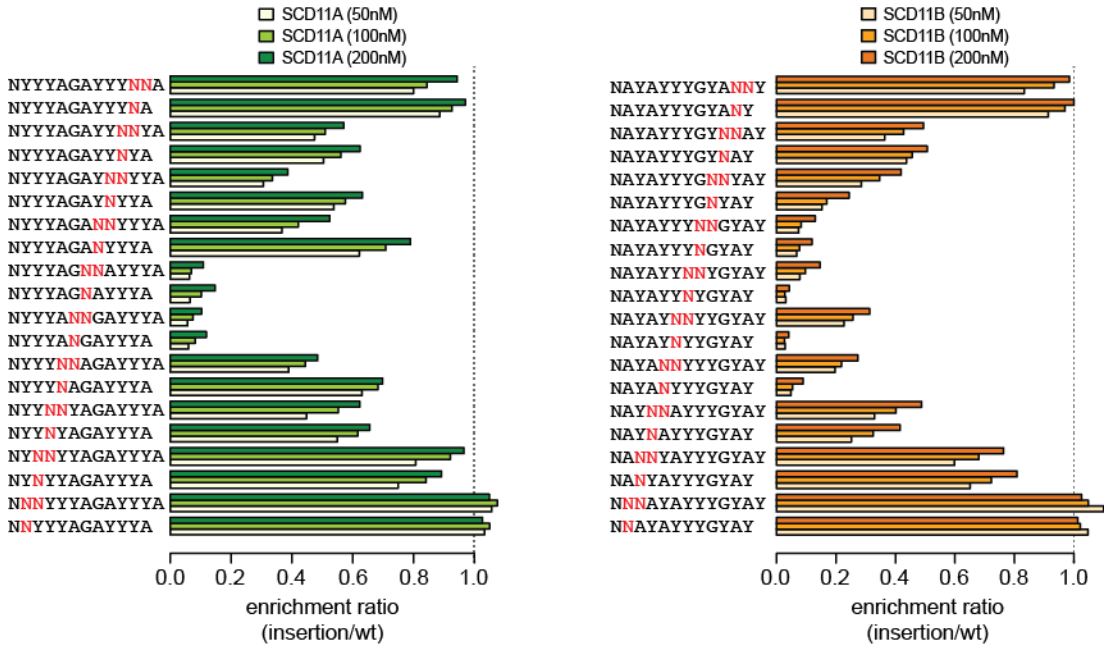
B



RNA	sequence
-2nt 5'	CUAGAUCUAGUU
14nt	GUCUAGAUCUAGUU
-2nt 3'	GUCUAGAUCUAG

Supplementary Figure S6

Supplementary Figure S7. Effects of small insertions on sequence enrichment by SCD11A and SCD11B. The analysis is the same as that in Figure 4 except that it displays the effects on sequence enrichment of insertions of one or two nucleotides at each position.



Supplementary Figure S7

REFERENCES

1. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, **31**, 3406-3415.
2. Pattanayak, V., Ramirez, C.L., Joung, J.K. and Liu, D.R. (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods*, **8**, 765-770.